PCT/US98/01065

02 FEBRUARY 1999 (02.02.99)

5

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

December 18, 1998

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/065,793***
FILING DATE: ***November 14, 1997***

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. Phillips

H. PHILLIPS
Certifying Officer

PTO/SB/16 (11/95) (Modified 1995 Legalsoft)
Approved for use through 01/31/98. OMB 0651-0037
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET
(Large Entity)

A/PROV

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (b)(2).

Docket Number	1038-766 MIS	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
PARRINGTON	Mark		45 Martin Street, Bradford, Ontario, Canada, L3Z 1Z4.

TITLE OF THE INVENTION (280 characters max)

ALPHAVIRUS VECTORS

CORRESPONDENCE ADDRESS

Michael I. Stewart
Sim & McBurney
6th Floor, 330 University Avenue
Toronto

STATE	Ontario	ZIP CODE	M5G 1R7	COUNTRY	Canada

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒	Specification	*Number of Pages*	8		
X	Drawing(s)	*Number of Sheets*	7	☐ Other (specify)	

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☒	A check or money order is enclosed to cover the filing fees	FILING FEE AMOUNT	$150.00
☐	The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number:		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number _____

Respectfully submitted,

SIGNATURE _____ Date 11/14/1997

TYPED or PRINTED NAME Michael I. Stewart REGISTRATION NO. *(if appropriate)* 24,973

☐ Additional inventors are being named on separately numbered sheets attached hereto

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, DC 20231

MIS 1038-766 1997 11 14 D1

TITLE OF INVENTION

ALPHAVIRUS VECTORS

FIELD OF INVENTION

The present invention relates to the field of DNA vaccines and is particularly concerned with modified alpha virus vectors for use in such vaccines.

BACKGROUND OF THE INVENTION

Semliki Forest virus (SFV) is a member of the Alphavirus genus in the Togaviridae family. The mature virus particle contains a single copy of a ssRNA genome with a positive polarity that is 5'-capped and 3'-polyadenylated. It functions as an mRNA and naked RNA can start an infection when introduced into cells. Upon infection/transfection, the 5' two-thirds of the genome is translated into a polyprotein that is processed into the four nonstructural proteins (nsP1 to 4) by self cleavage. Once the ns proteins have been synthesized they are responsible for replicating the plus-strand (42S) genome into full-length minus strands (ref. 14). These minus-strands then serve as templates for the synthesis of new plus-strand (42S) genomes and the 26S subgenomic mRNA (ref. 1 - Throughout this application, various references are cited in parentheses to describe more fully the state of the art to which this invention pertains. Full bibliographic information for each citation is found at the end of the specification. The disclosures of these references are hereby incorporated by reference into the present disclosure). This subgenomic mRNA, which is colinear with the last one-third of the genome, encodes the SFV structural proteins. In 1991 Liljestrom and Garoff (ref. 2) designed a series of expression vectors based on the SFV CDNA replicon. These vectors had the virus structural protein genes deleted to make the way for heterologous

inserts, but preserved the nonstructural coding region for production of the nsP1 to 4 replicase complex. Short 5' and 3' sequence elements required for RNA replication were also preserved. A polylinker site was inserted downstream from the 26S promoter followed by translation stop sites in all three frames. An SpeI site was inserted just after the 3' end of the SFV CDNA for linearization of the plasmid for use *in vitro* transcription reactions.

Injection of SFV RNA encoding a heterologous protein have been shown to result in the expression of the foreign protein and the induction of antibody in a number of studies (refs. 3,4). The use of SFV RNA inoculation to express foreign proteins for the purpose of immunization would have several of the advantages associated with plasmid DNA immunization. For example, SFV RNA encoding a viral antigen may be introduced in the presence of antibody to that virus without a loss in potency due to neutralization by antibodies to the virus. Also, because the protein is expressed *in vivo* the protein should have the same conformation as the protein expressed by the virus itself. Therefore, concerns about conformational changes which could occur during protein purification leading to a loss in immunogenicity, protective epitopes and possibly immunopotentiation, could be avoided by plasmid DNA immunization.

In WO95/27044, the disclosure of which is incorporated herein by reference, there is described the use of alphavirus cDNA vectors based on cDNA complementary to the alphavirus RNA sequence. Once transcribed from the cDNA under transceptional control of a heterologous promoter, the alphavirus RNA is able to self-replicate by means of its own replicase and thereby amplify the copy number of the transcribed recombinant RNA molecules.

SUMMARY OF THE INVENTION

The present invention is concerned with modifications to the alphavirus cDNA vectors described in the aforementioned WO 95/27044 to permit enhanced replication of the alphavirus. In the present invention, a heterologous splice site is introduced into the alphavirus replicon sequence, particularly that of Semliki Forest virus (SFV).

Accordingly, in one aspect, the present invention provides a cDNA molecule complementary to at least part of an alphavirus RNA genome, which cNDA molecule comprises the complement of the complete alphavirus RNA genome regions which are essential for replication of the said alphavirus RNA, and further comprises a heterologous cDNA sequence capable of expression in an animal or human host cell, said heterologous cDNA sequence being inserted into a region of the cDNA molecule which is non-essential to replication thereof, and the cDNA molecule being placed under transcriptional control of a promoter sequence functional in said animal or human cell, wherein at least one heterologous splice site is provided in the complement of the complete alphavirus RNA genome regions which are essential for replication of the alphavirus RNA, to prevent aberrant RNA splicing of the alphavirus.

The alphavirus molecule is a large molecule and, accordingly, there is a high probability of splice sites, thereby impairing the replication of the alphavirus and hence its ability to express the heterlogous DNA is impaired. By introducing the at least one heterologous splice site in accordance with the present invention into the alphavirus replicon sequence, any splicing is likely to be directed at the heterologous splice site rather than any endogenous splice site.

In the constructs provided herein, the promoter may be directly coupled to the 5'-end of the alphavirus sequence, which has the effect of reducing the variability in the splicing event at the 5'- end of the alphavirus replicon.

In addition, there may be provided at the 3'end of the Simliki Forest virus segment, a hepatitis delta ribosyme sequence to ensure proper in vivo cleavage at the 3'-end of the sequence. Any other convenient sequence may be employed to achieve this effect.

The heterologous splice site sequence may be provided by the nucleotide sequence of the rabbit β-globin intron II, as described in reference 5. Such heterologous splice site sequence may be inserted into the complement sequence at any convenient location which does not preclude replication of the alphavirus.

I have identified five suitable sites in the SFV replicon, which are contained within an EcoRV-SpeI fragment of the replicon which is 7983 bp in length (Fig. 3). The first such site is a Ppu-MI site, at position 2719 within the EcoRV-SpeI fragment.

In constructing the modified vectors provided herein, the EcoRV-SpeI fragment is cut with Ppu-MI at position 2719 and made blunt-ended with Mung Bean nuclease, which removes three bases from the SFV sequence. A blunt-ended β-globin II intron, which is 536 bp long, is ligated into the site and replaces the missing three bases with sequence added to the 3'-end of the β-globin intron sequence (Fig. 1).

The other four suitable sites for insertion of the Intron are the PvuII sites at bp 2518 3113, 6498 and 6872 of the EcoRV-SpeI fragment. Insertion of the Intron is achieved by cutting with PvuII (a blunt end cutter) and the blunt-ended β-globin II intron sequence (Fig. 2) is ligated into one or more of these sites.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the DNA sequence of the β-globin TT intron encoding three additional nucleotides at the 3'-end thereof (SEQ ID No:1);

Figure 2 shows the DNA sequence of the β-globin II intron (SEQ ID No:2); and

Figures 3A to 3E show the DNA sequence of the EcoRV-SpeI fragment of Semliki Forest virus replicon (SEQ ID No:3).

GENERAL DESCRIPTION OF INVENTION

As discussed above, the present invention provides a modified alphavirus cDNA. The alphavirus preferably is Semliki Forest virus.

The promoter sequence may comprise a promoter of eukaryotic or prokaryotic origin. Suitable promoters are the cytomegalovirus immediate early promoter (pCMV), althouqh other promoters, such as the Rous sarcoma virus long-terminal repeat promoter (pRSV), since, in the case of these and similar promoters, transcription is performed by the DNA-dependent RNA polymerase of the host cell. Additionally, the SP6, T3 or T7 promoters can be used, provided that the cell has first been transformed with genes encoding SP6, T3 or T7 RNA polymerase molecules which are either inserted into the chromosome or remain episomal. Expression of these (SP6, T3, T7) RNA polymerase-encoding genes is dependent on the host cell DNA-dependent RNA polymerase.

The heterologous cDNA insert may comprise the coding sequence for a desired product, which may be a biologically active protein or polypeptide, e.g., an immunogenic or antigenic protein or polypeptide, or a therapeutically active protein or polypeptide. The heterologous cDNA may also comprise additional

sequences, such as a sequence complementary to an RNA sequence which is a self-cleaving ribozyme sequence.

The DNA vectors provided herein may be administered to a host, including a human host, for *in vivo* expression of the heterologous cDNA sequence, in accordance with a further aspect of the invention, in order to generate an immune response in the host, which may be a protective immune response. The DNA vectors may be further formulated into immunogenic compositions for such administration.

SUMMARY OF DISCLOSURE

In summary of this disclosure, the present invention provides a modified alphavirus-based expression vector wherein at least one splice site is introduced to the alphavirus replicon to prevent aborrant splicing of the alphavirus genome. Modifications are possible within the scope of the invention.

REFERENCES

1. Fulginiti, V.A., Eller, J.J., Sieber, O.F., Joyner, J.W., Minamitani, M. and Meiklejohn, G., (1969) Am. J. Epidemiol. 89 (4), 435-448.

2. Chin, J., Magoffin, R.L., Shearer, L.A., Schieble, J.H. and Lennette, E.H. (1969) Am. J. Epidemiol. 89 (4), 449-463.

3. Jensen, K.E., Peeler, B.E. and Dulworth, W.G. (1962) J. Immunol. 89, 216-226.

4. Murphy, B.R., Prince, G.A., Collins, P.L., Van Wyke-Coelingh, K., Olmstead, R.A., Spriggs, M.K., Parrott, R.H., Kim, H.-Y., Brandt, C.D. and Chanock, R.N. (1988) Vir. Res. 11, 1-15.

5. Chapman, B.S.; Thayer, R.M.; Vincent, K.A. and Haigwood, N.L., Nucl. Acids. Res. 1991, 19: 3979-3986.

ABSTRACT OF THE DISCLOSURE

A modified alphavirus expression vector is provided wherein at least one heterologous splice site is introduced to the alphavirus replicon to prevent aborrant splicing of the alphavirus, which may be Semliki Forest virus following administration of the vector to a host.

```
        10          20          30          40          50          60
GTGAGTTTGG GGACCCTTGA TTGTTCTTTC TTTTTCGCTA TTGTAAAATT CATGTTATAT
        70          80          90         100         110         120
GGAGGGGGCA AAGTTTTCAG GGTGTTGTTT AGAATGGGAA GATGTCCCTT GTATCACCAT
       130         140         150         160         170         180
GGACCCTCAT GATAATTTTG TTTCTTTCAC TTTCTACTCT GTTGACAACC ATTGTCTCCT
       190         200         210         220         230         240
CTTATTTTCT TTTCATTTTC TGTAACTTTT TCGTTAAACT TTAGCTTGCA TTTGTAACGA
       250         260         270         280         290         300
ATTTTTAAAT TCACTTTTGT TTATTTGTCA GATTGTAAGT ACTTTCTCTA ATCACTTTTT
       310         320         330         340         350         360
TTTCAAGGCA ATCAGGGTAT ATTATATTGT ACTTCAGCAC AGTTTTAGAG AACAATTGTT
       370         380         390         400         410         420
ATAATTAAAT GATAAGGTAG AATATTTCTG CATATAAATT CTGGCTGGCG TGGAAATATT
       430         440         450         460         470         480
CTTATTGGTA GAAACAACTA CATCCTGGTC ATCATCCTGC CTTTCTCTTT ATGGTTACAA
       490         500         510         520         530         540
TGATATACAC TGTTTGAGAT GAGGATAAAA TACTCTGAGT CCAAACCGGG CCCCTCTGCT
       550         560         570         580         590         600
AACCATGTTC ATGCCTTCTT CTTTTTCCTA CAGGTC.... .......... ..........
```

B-globin +

3 SFV bases

FIG 1

```
        10          20          30          40          50          60
GTGAGTTTGG GGACCCTTGA TTGTTCTTTC TTTTTCGCTA TTGTAAAATT CATGTTATAT
        70          80          90         100         110         120
GGAGGGGGCA AAGTTTTCAG GGTGTTGTTT AGAATGGGAA GATGTCCCTT GTATCACCAT
       130         140         150         160         170         180
GGACCCTCAT GATAATTTTG TTTCTTTCAC TTTCTACTCT GTTGACAACC ATTGTCTCCT
       190         200         210         220         230         240
CTTATTTTCT TTTCATTTTC TGTAACTTTT TCGTTAAACT TTAGCTTGCA TTTGTAACGA
       250         260         270         280         290         300
ATTTTTAAAT TCACTTTTGT TTATTTGTCA GATTGTAAGT ACTTTCTCTA ATCACTTTTT
       310         320         330         340         350         360
TTTCAAGGCA ATCAGGGTAT ATTATATTGT ACTTCAGCAC AGTTTTAGAG AACAATTGTT
       370         380         390         400         410         420
ATAATTAAAT GATAAGGTAG AATATTTCTG CATATAAATT CTGGCTGGCG TGGAAATATT
       430         440         450         460         470         480
CTTATTGGTA GAAACAACTA CATCCTGGTC ATCATCCTGC CTTTCTCTTT ATGGTTACAA
       490         500         510         520         530         540
TGATATACAC TGTTTGAGAT GAGGATAAAA TACTCTGAGT CCAAACCGGG CCCCTCTGCT
       550         560         570         580         590         600
AACCATGTTC ATGCCTTCTT CTTTTTCCTA CAG..... .......... ..........
```

B-globin II

FIG2

EcoRV-SpeI Fragment

```
          10                  20                  30                  40                  50
ATCGGCAGTG CGCCTTCCAG GAGA[redacted]ATG TCTACGCACA AATACCACTG CGTATG[redacted]
          70                  80                  90                 100                 110                 120
ATGCGCAGCG CAGAAGACCC CGAAAGGCTC GATAGCTACG CAAAGAAACT GGCAGCGGCC
         130                 140                 150                 160                 170                 180
TCCGGGAAGG TGCTGGATAG AGAGATCGCA GGAAAAATCA CCGACCTGCA GACCGTCATG
         190                 200                 210                 220                 230                 240
GCTACGCCAG ACGCTGAATC TCCTACCTTT TGCCTGCATA CAGACGTCAC GTGTCGTACG
         250                 260                 270                 280                 290                 300
GCAGCCGAAG TGGCCGTATA CCAGGACGTG TATGCTGTAC ATGCACCAAC ATCGCTGTAC
         310                 320                 330                 340                 350                 360
CATCAGGCGA TGAAAGGTGT CAGAACGGCG TATTGGATTG GGTTTGACAC CACCCCGTTT
         370                 380                 390                 400                 410                 420
ATGTTTGACG CGCTAGCAGG CGCGTATCCA ACCTACGCCA CAAACTGGGC CGACGAGCAG
         430                 440                 450                 460                 470                 480
GTGTTACAGG CCAGGAACAT AGGACTGTGT GCAGCATCCT TGACTGAGGG AAGACTCGGC
         490                 500                 510                 520                 530                 540
AAACATGTCCA TTCTCCGCAA GAAGCAATTG AAACCTTGCG ACACAGTCAT GTTCTCGGTA
         550                 560                 570                 580                 590                 600
GGATCTACAT TGTACACTGA GAGCAGAAAG CTACTGAGGA GCTGGCACTT ACCCTCCGTA
         610                 620                 630                 640                 650                 660
TTCCACCTGA AAGGTAAACA ATCCTTTACC TGTAGGTGCG ATACCATCGT ATCATGTGAA
         670                 680                 690                 700                 710                 720
GGGTACGTAG TTAAGAAAAT CACTATGTGC CCCGGCCTGT ACGGTAAAAC GGTAGGGTAC
         730                 740                 750                 760                 770                 780
GCCGTGACGT ATCACGCGGA GGGATTCCTA GTGTGCAAGA CCACAGACAC TGTCAAAGGA
         790                 800                 810                 820                 830                 840
GAAAGAGTCT CATTCCCTGT ATGCACCTAC GTCCCCTCAA CCATCTGTGA TCAAATGACT
         850                 860                 870                 880                 890                 900
GGCATACTAG CGACCGACGT CACACCGGAG GACGCACAGA AGTTGTTAGT GGGATTGAAT
         910                 920                 930                 940                 950                 960
CAGAGGATAG TTGTGAACGG AAGAACACAG CGAAACACTA ACACGATGAA GAACTATCTG
         970                 980                 990                1000                1010                1020
CTTCCGATTG TGGCCGTCGC ATTTAGCAAG TGGGCGAGGG AATACAAGGC AGACCTTGAT
        1030                1040                1050                1060                1070                1080
GATGAAAAAC CTCTGGGTGT CCGAGAGAGG TCACTTACTT GCTGCTGCTT GTGGGCATTT
        1090                1100                1110                1120                1130                1140
AAAACGAGGA AGATGCACAC CATGTACAAG AAACCAGACA CCCAGACAAT AGTGAAGGTG
        1150                1160                1170                1180                1190                1200
CCTTCAGAGT TTAACTCGTT CGTCATCCCG AGCCTATGGT CTACAGGCCT CGCAATCCCA
        1210                1220                1230                1240                1250                1260
GTCAGATCAC GCATTAAGAT GCTTTTGGCC AAGAAGACCA AGCGAGAGTT AATACCTGTT
        1270                1280                1290                1300                1310                1320
CTCGACGCGT CGTCAGCCAG GGATGCTGAA CAAGAGGAGA AGGAGAGGTT GGAGGCCGAG
        1330                1340                1350                1360                1370                1380
CTGACTAGAG AAGCCTTACC ACCCCTCGTC CCCATCGCGC CGGCGGAGAC GGGAGTCGTC
        1390                1400                1410                1420                1430                1440
GACGTCGACG TTGAAGAACT AGAGTATCAC GCAGGTGCAG GGGTCGTGGA AACACCTCGC
        1450                1460                1470                1480                1490                1500
AGCGCGTTGA AAGTCACCGC ACAGCCGAAC GACGTACTAC TAGGAAATTA CGTAGTTCTG
        1510                1520                1530                1540                1550                1560
TCCCCGCAGA CCGTGCTCAA GAGCTCCAAG TTGGCCCCCG TGCACCCTCT AGCAGAGCAG
        1570                1580                1590                1600                1610                1620
GTGAAAATAA TAACACATAA CGGGAGGGCC GGCGGTTACC AGGTCGACGG ATATGACGGC
        1630                1640                1650                1660                1670                1680
AGGGTCCTAC TACCATGTGG ATCGGCCATT CCGGTCCCTG AGTTTCAAGC TTTGAGCGAG
        1690                1700                1710                1720                1730                1740
AGCGCCACTA TGGTGTACAA CGAAAGGGAG TTCGTCAACA GGAAACTATA CCATATTGCC
        1750                1760                1770                1780                1790                1800
```


FIG 3A

```
GTTCACGGAC CGTCGCTGAA CA[redacted]ACGAG GAGAACTACG AGAAAGTCAG AGCT[redacted]GA
      1810         1820         1830         1840         1850         1860
ACTGACGCCG AGTACGTGTT CGACGTAGAT AAAAAATGCT GCGTCAAGAG AGAGGAAGCG
      1870         1880         1890         1900         1910         1920
TCGGGTTTGG TGTTGGTGGG AGAGCTAACC AACCCCCCGT TCCATGAATT CGCCTACGAA
      1930         1940         1950         1960         1970         1980
GGGCTGAAGA TCAGGCCGTC GGCACCATAT AAGACTACAG TAGTAGGAGT CTTTGGGGTT
      1990         2000         2010         2020         2030         2040
CCGGGATCAG GCAAGTCTGC TATTATTAAG AGCCTCGTGA CCAAACACGA TCTGGTCACC
      2050         2060         2070         2080         2090         2100
AGCGGCAAGA AGGAGAACTG CCAGGAAATA GTTAACGACG TGAAGAAGCA CCGCGGGAAG
      2110         2120         2130         2140         2150         2160
GGGACAAGTA GGGAAAACAG TGACTCCATC CTGCTAAACG GGTGTCGTCG TGCCGTGGAC
      2170         2180         2190         2200         2210         2220
ATCCTATATG TGGACGAGGC TTTCGCTTGC CATTCCGGTA CTCTGCTGGC CCTAATTGCT
      2230         2240         2250         2260         2270         2280
CTTGTTAAAC CTCGGAGCAA AGTGGTGTTA TGCGGAGACC CCAAGCAATG CGGATTCTTC
      2290         2300         2310         2320         2330         2340
AATATGATGC AGCTTAAGGT GAACTTCAAC CACAACATCT GCACTGAAGT ATGTCATAAA
      2350         2360         2370         2380         2390         2400
AGTATATCCA GACGTTGCAC GCGTCCAGTC ACGGCCATCG TGTCTACGTT GCACTACGGA
      2410         2420         2430         2440         2450         2460
GGCAAGATGC GCACGACCAA CCCGTGCAAC AAACCCATAA TCATAGACAC CACAGGACAG
      2470         2480         2490         2500         2510         2520
ACCAAGCCCA AGCCAGGAGA CATCGTGTTA ACATGCTTCC GAGGCTGGGC AAAGCAGCTG
      2530         2540         2550         2560         2570         2580
CAGTTGGACT ACCGTGGACA CGAAGTCATG ACAGCAGCAG CATCTCAGGG CCTCACCCGC
      2590         2600         2610         2620         2630         2640
AAAGGGGTAT ACGCCGTAAG GCAGAAGGTG AATGAAAATC CCTTGTATGC CCCTGCGTCG
      2650         2660         2670         2680         2690         2700
GAGCACGTGA ATGTACTGCT GACGCGCACT GAGGATAGGC TGGTGTGGAA AACGCTGGCC
      2710         2720         2730         2740         2750         2760
GGCGATCCCT GGATTAAGGT CCTATCAAAC ATTCCACAGG GTAACTTTAC GGCCACATTG
      2770         2780         2790         2800         2810         2820
GAAGAATGGC AAGAAGAACA CGACAAAATA ATGAAGGTGA TTGAAGGACC GGCTGCGCCT
      2830         2840         2850         2860         2870         2880
GTGGACGCGT TCCAGAACAA AGCGAACGTG TGTTGGGCGA AAAGCCTGGT GCCTGTCCTG
      2890         2900         2910         2920         2930         2940
GAGACTGCCG GAATCAGATT GACAGCAGAG GAGTGGAGCA CCATAATTAC AGCATTTAAG
      2950         2960         2970         2980         2990         3000
GAGGACAGAG CTTACTCTCC AGTGGTGGCC TTGAATGAAA TTTGCACCAA GTACTATGGA
      3010         3020         3030         3040         3050         3060
GTTGACCTGG ACAGTGGCCT GTTTTCTGCC CCGAAGGTGT CCCTGTATTA CGAGAACAAC
      3070         3080         3090         3100         3110         3120
CACTGGGATA ACAGACCTGG TGGAAGGATG TATGGATTCA ATGCCGCAAC AGCTGCCAGG
      3130         3140         3150         3160         3170         3180
CTGGAAGCTA GACATACCTT CCTGAAGGGG CAGTGGCATA CGGGCAAGCA GGCAGTTATC
      3190         3200         3210         3220         3230         3240
GCAGAAAGAA AAATCCAACC GCTTTCTGTG CTGGACAATG TAATTCCTAT CAACCGCAGG
      3250         3260         3270         3280         3290         3300
CTGCCGCACG CCCTGGTGGC TGAGTACAAG ACGGTTAAAG GCAGTAGGGT TGAGTGGCTG
      3310         3320         3330         3340         3350         3360
GTCAATAAAG TAAGAGGGTA CCACGTCCTG CTGGTGAGTG AGTACAACCT GGCTTTGCCT
      3370         3380         3390         3400         3410         3420
CGACGCAGGG TCACTTGGTT GTCACCGCTG AATGTCACAG GCGCCGATAG GTGCTACGAC
      3430         3440         3450         3460         3470         3480
CTAAGTTTAG GACTGCCGGC TGACGCCGGC AGGTTCGACT TGGTCTTTGT GAACATTCAC
      3490         3500         3510         3520         3530         3540
ACGGAATTCA GAATCCACCA CTACCAGCAG TGTGTCGACC ACGCCATGAA GCTGCAGATG
```


FIG 3B

```
      3550        3560        3570        3580        3590        3600
CTTGGGGGAG ATGCGCTACG ACTGG[redacted]AA CCCGGCGGCA TCTTGATGAG AGCTTAC[redacted]
      3610        3620        3630        3640        3650        3660
TACGCCGATA AAATCAGCGA AGCCGTTGTT TCCTCCTTAA GCAGAAAGTT CTCGTCTGCA
      3670        3680        3690        3700        3710        3720
AGAGTGTTGC GCCCGGATTG TGTCACCAGC AATACAGAAG TGTTCTTGCT GTTCTCCAAC
      3730        3740        3750        3760        3770        3780
TTTGACAACG GAAAGAGACC CTCTACGCTA CACCAGATGA ATACCAAGCT GAGTGCCGTG
      3790        3800        3810        3820        3830        3840
TATGCCGGAG AAGCCATGCA CACGGCCGGG TGTGCACCAT CCTACAGAGT TAAGAGAGCA
      3850        3860        3870        3880        3890        3900
GACATAGCCA CGTGCACAGA AGCGGCTGTG GTTAACGCAG CTAACGCCCG TGGAACTGTA
      3910        3920        3930        3940        3950        3960
GGGGATGGCG TATGCAGGGC CGTGGCGAAG AAATGGCCGT CAGCCTTTAA GGGAGCAGCA
      3970        3980        3990        4000        4010        4020
ACACCAGTGG GCACAATTAA AACAGTCATG TGCGGCTCGT ACCCCGTCAT CCACGCTGTA
      4030        4040        4050        4060        4070        4080
GCGCCTAATT TCTCTGCCAC GACTGAAGCG GAAGGGGACC GCGAATTGGC CGCTGTCTAC
      4090        4100        4110        4120        4130        4140
CGGGCAGTGG CCGCCGAAGT AAACAGACTG TCACTGAGCA GCGTAGCCAT CCCGCTGCTG
      4150        4160        4170        4180        4190        4200
TCCACAGGAG TGTTCAGCGG CGGAAGAGAT AGGCTGCAGC AATCCCTCAA CCATCTATTC
      4210        4220        4230        4240        4250        4260
ACAGCAATGG ACGCCACGGA CGCTGACGTG ACCATCTACT GCAGAGACAA AAGTTGGGAG
      4270        4280        4290        4300        4310        4320
AAGAAAATCC AGGAAGCCAT TGACATGAGG ACGGCTGTGG AGTTGCTCAA TGATGACGTG
      4330        4340        4350        4360        4370        4380
GAGCTGACCA CAGACTTGGT GAGAGTGCAC CCGGACAGCA GCCTGGTGGG TCGTAAGGGC
      4390        4400        4410        4420        4430        4440
TACAGTACCA CTGACGGGTC GCTGTACTCG TACTTTGAAG GTACGAAATT CAACCAGGCT
      4450        4460        4470        4480        4490        4500
GCTATTGATA TGGCAGAGAT ACTGACGTTG TGGCCCAGAC TGCAAGAGGC AAACGAACAG
      4510        4520        4530        4540        4550        4560
ATATGCCTAT ACGCGCTGGG CGAAACAATG GACAACATCA GATCCAAATG TCCGGTGAAC
      4570        4580        4590        4600        4610        4620
GATTCCGATT CATCAACACC TCCCAGGACA GTGCCCTGCC TGTGCCGCTA CGCAATGACA
      4630        4640        4650        4660        4670        4680
GCAGAACGGA TCGCCCGCCT TAGGTCACAC CAAGTTAAAA GCATGGTGGT TTGCTCATCT
      4690        4700        4710        4720        4730        4740
TTTCCCCTCC CGAAATACCA TGTAGATGGG GTGCAGAAGG TAAAGTGCGA GAAGGTTCTC
      4750        4760        4770        4780        4790        4800
CTGTTCGACC CGACGGTACC TTCAGTGGTT AGTCCGCGGA AGTATGCCGC ATCTACGACG
      4810        4820        4830        4840        4850        4860
GACCACTCAG ATCGGTCGTT ACGAGGGTTT GACTTGGACT GGACCACCGA CTCGTCTTCC
      4870        4880        4890        4900        4910        4920
ACTGCCAGCG ATACCATGTC GCTACCCAGT TTGCAGTCGT GTGACATCGA CTCGATCTAC
      4930        4940        4950        4960        4970        4980
GAGCCAATGG CTCCCATAGT AGTGACGGCT GACGTACACC CTGAACCCGC AGGCATCGCG
      4990        5000        5010        5020        5030        5040
GACCTGGCGG CAGATGTGCA CCCTGAACCC GCAGACCATG TGGACCTCGA GAACCCGATT
      5050        5060        5070        5080        5090        5100
CCTCCACCGC GCCCGAAGAG AGCTGCATAC CTTGCCTCCC GCGCGGCGGA GCGACCGGTG
      5110        5120        5130        5140        5150        5160
CCGGCGCCGA GAAAGCCGAC GCCTGCCCCA AGGACTGCGT TTAGGAACAA GCTGCCTTTG
      5170        5180        5190        5200        5210        5220
ACGTTCGGCG ACTTTGACGA GCACGAGGTC GATGCGTTGG CCTCCGGGAT TACTTTCGGA
      5230        5240        5250        5260        5270        5280
GACTTCGACG ACGTCCTGCC ACTAGGCCGC GCGGGTGCAT ATATTTTCTC CTCGGACACT
      5290        5300        5310        5320        5330        5340
```


FIG 3C

```
GGCAGCGGAC ATTTACAACA AA[redacted]TCGTT AGGCAGCACA ATCTCCAGTG CGCA[redacted]TG
      5350       5360       5370       5380       5390       5400
GATGCGGTCC AGGAGGAGAA AATGTACCCG CCAAAATTGG ATACTGAGAG GGAGAAGCTG
      5410       5420       5430       5440       5450       5460
TTGCTGCTGA AAATGCAGAT GCACCCATCG GAGGCTAATA AGAGTCGATA CCAGTCTCGC
      5470       5480       5490       5500       5510       5520
AAAGTGGAGA ACATGAAAGC CACGGTGGTG GACAGGCTCA CATCGGGGGC CAGATTGTAC
      5530       5540       5550       5560       5570       5580
ACGGGAGCGG ACGTAGGCCG CATACCAACA TACGCGGTTC GGTACCCCCG CCCCGTGTAC
      5590       5600       5610       5620       5630       5640
TCCCCTACCG TGATCGAAAG ATTCTCAAGC CCCGATGTAG CAATCGCAGC GTGCAACGAA
      5650       5660       5670       5680       5690       5700
TACCTATCCA GAAATTACCC AACAGTGGCG TCGTACCAGA TAACAGATGA ATACGACGCA
      5710       5720       5730       5740       5750       5760
TACTTGGACA TGGTTGACGG GTCGGATAGT TGCTTGGACA GAGCGACATT CTGCCCGGCG
      5770       5780       5790       5800       5810       5820
AAGCTCCGGT GCTACCCGAA ACATCATGCG TACCACCAGC CGACTGTACG CAGTGCCGTC
      5830       5840       5850       5860       5870       5880
CCGTCACCCT TTCAGAACAC ACTACAGAAC GTGCTAGCGG CCGCCACCAA GAGAAACTGC
      5890       5900       5910       5920       5930       5940
AACGTCACGC AAATGCGAGA ACTACCCACC ATGGACTCGG CAGTGTTCAA CGTGGAGTGC
      5950       5960       5970       5980       5990       6000
TTCAAGCGCT ATGCCTGCTC CGGAGAATAT TGGGAAGAAT ATGCTAAACA ACCTATCCGG
      6010       6020       6030       6040       6050       6060
ATAACCACTG AGAACATCAC TACCTATGTG ACCAAATTGA AAGGCCCGAA AGCTGCTGCC
      6070       6080       6090       6100       6110       6120
TTGTTCGCTA AGACCCACAA CTTGGTTCCG CTGCAGGAGG TTCCCATGGA CAGATTCACG
      6130       6140       6150       6160       6170       6180
GTCGACATGA AACGAGATGT CAAAGTCACT CCAGGGACGA AACACACAGA GGAAAGACCC
      6190       6200       6210       6220       6230       6240
AAAGTCCAGG TAATTCAAGC AGCGGAGCCA TTGGCGACCG CTTACCTGTG CGGCATCCAC
      6250       6260       6270       6280       6290       6300
AGGGAATTAG TAAGGAGACT AAATGCTGTG TTACGCCCTA ACGTGCACAC ATTGTTTGAT
      6310       6320       6330       6340       6350       6360
ATGTCGGCCG AAGACTTTGA CGCGATCATC GCCTCTCACT TCCACCCAGG AGACCCGGTT
      6370       6380       6390       6400       6410       6420
CTAGAGACGG ACATTGCATC ATTCGACAAA AGCCAGGACG ACTCCTTGGC TCTTACAGGT
      6430       6440       6450       6460       6470       6480
TTAATGATCC TCGAAGATCT AGGGGTGGAT CAGTACCTGC TGGACTTGAT CGAGGCAGCC
      6490       6500       6510       6520       6530       6540
TTTGGGGAAA TATCCAGCTG TCACCTACCA ACTGGCACGC GCTTCAAGTT CGGAGCTATG
      6550       6560       6570       6580       6590       6600
ATGAAATCGG GCATGTTTCT GACTTTGTTT ATTAACACTG TTTTGAACAT CACCATAGCA
      6610       6620       6630       6640       6650       6660
AGCAGGGTAC TGGAGCAGAG ACTCACTGAC TCCGCCTGTG CGGCCTTCAT CGGCGACGAC
      6670       6680       6690       6700       6710       6720
AACATCGTTC ACGGAGTGAT CTCCGACAAG CTGATGGCGG AGAGGTGCGC GTCGTGGGTC
      6730       6740       6750       6760       6770       6780
AACATGGAGG TGAAGATCAT TGACGCTGTC ATGGGCGAAA AACCCCCATA TTTTTGTGGG
      6790       6800       6810       6820       6830       6840
GGATTCATAG TTTTTGACAG CGTCACACAG ACCGCCTGCC GTGTTTCAGA CCCACTTAAG
      6850       6860       6870       6880       6890       6900
CGCCTGTTCA AGTTGGGTAA GCCGCTAACA GCTGAAGACA AGCAGGACGA AGACAGGCGA
      6910       6920       6930       6940       6950       6960
CGAGCACTGA GTGACGAGGT TAGCAAGTGG TTCCGGACAG GCTTGGGGGC CGAACTGGAG
      6970       6980       6990       7000       7010       7020
GTGGCACTAA CATCTAGGTA TGAGGTAGAG GGCTGCAAAA GTATCCTCAT AGCCATGGCC
      7030       7040       7050       7060       7070       7080
ACCTTGGCGA GGGACATTAA GGCGTTTAAG AAATTGAGAG GACCTGTTAT ACACCTCTAC
```


FIG 3D

```
      7090        7100                    7120        7130        714
GGCGGTCCTA GATTGGTGCG TTAATACGCA GAATTCTGAT TGGATCCCGG GTAATTAA
      7150        7160        7170        7180        7190        7200
GAATTACATC CCTACGCAAA CGTTTTACGG CCGCCGGTGG CGCCCGCGCC CGGCGGCCCG
      7210        7220        7230        7240        7250        7260
TCCTTGGCCG TTGCAGGCCA CTCCGGTGGC TCCCGTCGTC CCCGACTTCC AGGCCCAGCA
      7270        7280        7290        7300        7310        7320
GATGCAGCAA CTCATCAGCG CCGTAAATGC GCTGACAATG AGACAGAACG CAATTGCTCC
      7330        7340        7350        7360        7370        7380
TGCTAGGCCT CCCAAACCAA AGAAGAAGAA GACAACCAAA CCAAAGCCGA AAACGCAGCC
      7390        7400        7410        7420        7430        7440
CAAGAAGATC AACGGAAAAA CGCAGCAGCA AAAGAAGAAA GACAAGCAAG CCGACAAGAA
      7450        7460        7470        7480        7490        7500
GAAGAAGAAA CCCGGAAAAA GAGAAAGAAT GTGCATGAAG ATTGAAAATG ACTGTATCTT
      7510        7520        7530        7540        7550        7560
CGTATGCGGC TAGCCACAGT AACGTAGTGT TTCCAGACAT GTCGGGCACC GCACTATCAT
      7570        7580        7590        7600        7610        7620
GGGTGCAGAA AATCTCGGGT GGTCTGGGGG CCTTCGCAAT CGGCGCTATC CTGGTGCTGG
      7630        7640        7650        7660        7670        7680
TTGTGGTCAC TTGCATTGGG CTCCGCAGAT AAGTTAGGGT AGGCAATGGC ATTCATATAG
      7690        7700        7710        7720        7730        7740
CAAGAAAATT GAAAACAGAA AAAGTTAGGG TAAGCAATGG CATATAACCA TAACTGTATA
      7750        7760        7770        7780        7790        7800
ACTTGTAACA AAGCGCAACA AGACCTGCGC AATTGGCCCC GTGGTCCGCC TCACGGAAAC
      7810        7820        7830        7840        7850        7860
TCGGGGCAAC TCATATTGAC ACATTAATTG GCAATAATTG GAAGCTTACA TAAGCTTAAT
      7870        7880        7890        7900        7910        7920
TGGACGAATA ATTGGATTTT TATTTTATTT TGCAATTGGT TTTTAATATT TCCAAAAAAA
      7930        7940        7950        7960        7970        7980
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
      7990        8000        8010        8020        8030        8040
AAA.......  ..........  ..........  ..........  ..........  ..........
```


FIG 3E